Toward Silent–Speech Control of Consumer Wearables

Abdelkareem Bedri, Himanshu Sahni, Pavleen Thukral, Thad Starner, David Byrd, Peter Presti, Gabriel Reyes, and Maysam Ghovanloo, Georgia Tech

Zehua Guo, Microsoft

Systems that recognize silent speech can enable fast, hands–free communication. Two prototypes let users control Google Glass with tongue movements and jaw gestures, requiring no additional equipment except a tongue–mounted magnet or consumer earphones augmented with embedded proximity sensors.

Loss of intelligible speech caused by neurological injury of the motor-speech system's motor component (dysarthria) or the inability to produce voice (aphonia) can be a significant barrier to communication with family, friends, and caregivers. Augmentative and alternative communication (AAC) aids can assist with communication, but these systems tend to be significantly slower than speech. For individuals rendered incapable of intelligible speech but who retain good mouth articulation or normal tongue movement capabilities, silent-speech recognition systems can allow communication at a faster pace.

For firefighters, aviators, combat soldiers, or other individuals in environments with high ambient noise levels, silent-speech recognition systems can enable fast, hands-free control as users interact with computing devices. In some situations, even simple jaw gestures might be sufficient for communication. For example, a special operations soldier whose mission requires silence could acknowledge a radio message by wiggling his jaw from side to side.

Even consumer Bluetooth headset users might find it beneficial to have an interface that recognizes silent speech and jaw gestures. During a meeting or class, the user might silently mouth "send to voicemail" to redirect an incoming phone call and avoid distracting nearby colleagues.

A key requirement for these applications is their ability to be unobtrusive, even hard to differentiate from mainstream consumer devices such as headphones. Recognizing this need, we developed and experimented with two systems for silent-speech recognition that work with Google Glass and thus could be made indistinguishable from off-the-shelf devices. The tongue magnet interface (TMI) is designed for Glass users who would be willing to have a magnet placed temporarily on the tongue or to have a magnetic tongue piercing. For other users, we created the outer ear interface (OEI), which also detects silent speech and jaw movements but can be incorporated into a common consumer audio headset.

TABLE 1. Summary of experiments and results.

Experiment	No. of subjects	Parameters	Recognition technique	Overall or average accuracy (%)
Feasibility (tongue drive system)	4	12 phrases, 1,172 samples	Hidden Markov models (HMMs)	96.00
Tongue magnet interface + outer ear interface (OEI)	6	11 phrases, 1,901 samples	HMMs	90.50
OEI alone	1	9 phrases, 225 samples	HMMs	85.34
Improved OEI with simple jaw gesture	22	6 classes, 1,584 samples	Dynamic time warping	User dependent: 97.64 User independent: 84.51

We conducted a series of experiments, listed in Table 1, to explore our systems' capabilities.[1] We first performed a feasibility experiment using the tongue drive system (TDS), custom hardware developed at Georgia Tech.[2] Once we established the feasibility of using the TDS for silent-speech recognition, we began experimenting with the TMI, first alone and then with a single proximity sensor version of the OEI in an ear mold. The combined TMI and OEI system employed user-dependent model training, and average recognition accuracy of a set of 11 phrases was as high as 90.5 percent. To develop a system that could appeal to more users and be perceived as normal consumer electronics, we housed three embedded proximity sensors in a Sony jogging headset.

Although we used an external laptop for computation in our experiments, Glass is sufficiently powerful to run the recognition algorithms on-board as a self-contained system, which eliminates the need for additional processing equipment.

FEASIBILITY EXPERIMENT WITH THE TDS

The TDS is a custom-manufactured mechanism designed to help paralyzed users control a power wheelchair with their tongue. Figure 1 shows the TDS headset, which contains four three-axis HMC1043 Honeywell magnetometers mounted on bendable gooseneck tubes. The TDS tracks the 3D position of an encapsulated rare-earth magnet in a tongue stud or glued about five millimeters from the tip of the user's tongue.[1]

In a typical TDS application, the user touches his tongue to different teeth to turn the wheelchair or move it forward or backward. However, because our goal was to detect relative tongue movements when the user mouths a phrase, we were less concerned with the exact position of the magnet. Consequently, we focused on the magnet's relative motion as reflected by a change in magnetic field strength.

Test phrases

The feasibility test involved four subjects, from whom we collected 25 samples of 12 phrases each. The phrases, which are phonetically distinct and reflect our intent to use the system as a communication aid, were as follows:

Notice the context of all these stmts
- I want water.
- Food please, I'm hungry.
- It hurts.
- How are you today?
- The tongue magnet interface is spectacular.
- Thank you, I'm great.
- Give me my medicine, please.
- Thanks for watering my plants.
- My mother is hurt and wants medicine.
- I need to use the bathroom.
- The medicine hurt me.
- Help me, I need assistance.

To collect the recordings needed to train the phrase-recognition models, we displayed the phrases in a sequence at regular intervals with breaks every 10 samples, randomizing the sequence to avoid data-collection bias that can arise from fatigue and gradual adjustment to the hardware. At each prompt, the subject silently spoke the sample phrase in a natural way, raising a hand when the phrase was complete. We encouraged subjects to keep their heads as steady as possible during recording, but otherwise placed no restrictions on their speech.

Data preprocessing

We used a Gaussian filter to smooth the data and then segmented it into individual samples and labeled them. For segmentation, we used an automatic sliding-window, variance-based scheme to detect silences and separate them from utterances, which yielded a 12-dimensional time series (from the headset's four three-axis magnetometers). We applied principal component analysis to reduce the number of dimensions, which allowed us to estimate fewer parameters. This process, combined with the TDS's redundant sensors, helps reduce noise. Indeed, the first three principal
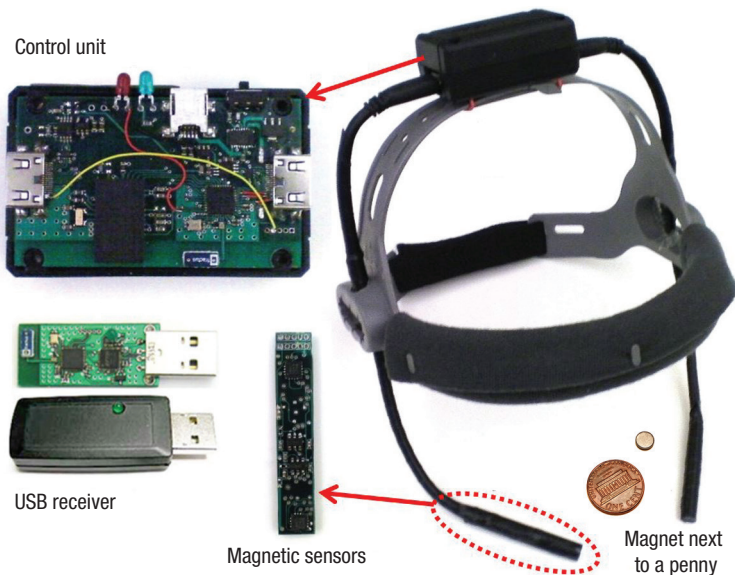
FIGURE 1. The tongue drive system (TDS) used to determine the feasibility of silent speech. The user wears a headset along with a 0.5-mm magnet adhered to the tongue. When the tongue moves, magnetic sensors identify a change in magnetic field strength, which the headset communicates wirelessly to a desktop using a 2.4-GHz industrial, scientific, and medical (ISM) band transceiver. (Headset, control unit, receiver, and sensors images from J. Kim, "Evaluation of a Smartphone Platform as a Wireless Interface between Tongue Drive System and Electric-Powered Wheelchairs," *IEEE Trans. Biomedical Eng.*, vol. 59, no. 6, 2012, pp. 1787–1796.)

components accounted for an average 80.44 percent of headset data variance.

Of the total 1,200 phrase samples collected, we discarded 28 phrase segments because of sensor errors, such as shot noise and wireless packet dropping.

Recognition technique

To recognize phrases, we used hidden Markov models (HMMs) and the HTK 3.4.1 toolkit,[3] using one of three empirically derived topologies, depending on utterance length. We used a three-state, one-skip topology at the start and end of each phrase and represented all states with diagonal covariance matrices.

- For phrases with fewer than nine phonemes, we used a six-state, left-right topology with two skip transitions (state 1 → state 3 and state 3 → state 5).
- For phrases with 10 to 18 phonemes, we used a nine-state, left-right, no-skip topology.
- For phrases with more than 18 phonemes, we used a 12-state, left-right, no-skip topology.

For testing, we separated each user's data randomly into mutually exclusive sets: training (70 percent of the data) and test (30 percent of the data). Thus, the system is user dependent—trained for each user. We performed 20-fold Monte Carlo cross-validation to determine the average accuracy for each user. Across the four users, the TDS implementation achieved 96 percent average recognition accuracy.

INTERFACE CONSTRUCTION

Encouraged by the feasibility experiment results, we envisioned creating a smaller TDS that could integrate with Glass. However, we first wanted to try building an interface that would not require modifying Glass but instead could use its internal three-axis magnetometer mounted on the front of the device close to the display.

Our initial efforts produced the TMI, which uses Glass's magnetometer to measure the movement of the magnet affixed to the user's tongue. Shortly after, we developed the OEI, which slightly extends Glass with earphones

that use embedded proximity sensors to measure the ear and ear-canal deformation caused by jaw movements. Figure 2 shows how the TMI and OEI prototypes integrate with Glass.

A magnet temporarily glued to the tongue is awkward to manage, but piercing the tongue to retain the magnet permanently—an alternative that some TDS users have adopted—might seem too extreme for Glass users other than those with vocal disabilities or special vocational needs. Consequently, while we conducted experiments with the TMI prototype, we also investigated other ways of recognizing silent speech.

The volume of work on silent speech, some of which is briefly described in the sidebar "Work on Silent-Speech Interfaces," has established that jaw motion is an important part of speech production. Some of this work motivated us to create the OEI, which detects lower-jaw (mandible) movement by measuring the deformation caused in the ear canal.[4]

As Figures 3a and 3b illustrate, the OEI's proximity sensors—in earpieces at the canal's outer edge, one for each ear—measure the mandibular condyle's location from the volume of expansion in the ear canal. As the jaw moves during silent speech production, the distance to the ear canal wall changes, which the proximity sensors register.

Our experiments showed that the sensors were sufficiently unobtrusive that we were able to later integrate them into standard Sony sports earphones.

TMI + OEI EXPERIMENT

To determine what our interfaces could detect, we combined their use in an experiment to recognize silently spoken phrases. The OEI used a Teensy 3.1-based data-acquisition unit to gather

WORK ON SILENT-SPEECH INTERFACES

Silent-speech recognition systems use mouth movements to speed communication through augmentative and alternative communication devices and to control automated systems that help individuals with certain disabilities remain independent.[1] Among the seven technology categories defined for such interfaces,[1] electromagnetic articulography (EMA) sensors have the most relevance to a system that reads magnetic signals. EMA sensors are inexpensive and excellent for silent-speech recognition and speech recognition in noisy environments. However, most EMA interfaces are not straightforward—some consist of externally excited, implanted coils with wires that users find extremely inconvenient—and thus face low user acceptance. In contrast, our tongue magnet interface (TMI) allows users to wear a mainstream consumer device, such as Google Glass, and requires only one magnet.

MOUTH INTERFACES

A system similar to the TMI uses seven passive magnets attached to the lips, teeth, and tongue, sensed by six dual-axis magnetometers mounted in a pair of glasses.[2] Using a dynamic time warping (DTW) algorithm with a modified distance function, the system achieved 97 percent accuracy on nine isolated words and 94 percent accuracy on 13 isolated phonemes. Although this apparatus is promising, it is awkward to wear and maintain and has yet to address phrase-level communication.

Another approach distinguishes five tongue gestures, which the user executes by pressing his tongue on the inside of his cheek.[3] An array of textile pressure sensors on the outer cheek captures pressure intensity and tongue movement. Using k nearest neighbors, the system achieved 80 percent accuracy per frame and 98 percent accuracy by voting across the interaction interval. Despite its high accuracy, the system suffers from poor durability and social acceptance.

EAR INTERFACES

Another study shows the feasibility of distinguishing tongue gestures using only a microphone in the ear canal.[4] It is similar to our outer ear interface (OEI) in that it senses jaw movement from the ear canal. The system achieved more than 97 percent accuracy in classifying inner-ear pressure signals as one of four tongue motions, and its creators suggest that tongue movements can be useful in hands-free communication and control applications—further evidence that more complex tongue motions, such as those involved in speech, can be recognizable though an ear sensor.

The Heartphone system uses reflective photo sensors embedded in a pair of ear buds.[5] The system detects the user's average heart rate by measuring the amount of reflected light from volumetric changes in the blood vessels located in the tragus of a human body at rest. Heartphone has an average error of 0.63 percent.

Another ear interface for gesture recognition, the in-ear biofeed controller, uses a gyroscope and physiological sensors integrated in an earphone.[6] After processing the data captured by the two sensors, the system can classify gestures such as head nodding and shaking, eye winking, and ear wiggling, which can serve as control input to a mobile device.

Electrooculography was the basis for several prototype ear pieces containing three electrodes that capture eye gestures.[7] However, the error rate was strongly subject dependent.

References

1. B. Denby et al., "Silent Speech Interfaces," *Speech Communication*, vol. 52, no. 4, 2010, pp. 270–287.

2. M.J. Fagan et al., "Development of a (Silent) Speech Recognition System for Patients Following Laryngectomy," *Medical Eng. & Physics*, vol. 30, no. 4, 2008, pp. 419–425.

3. J. Cheng et al., "On the Tip of My Tongue: A Non-Invasive Pressure-Based Tongue Interface," *Proc. 5th ACM Int'l Conf. Augmented Human* (AH 14), 2014, article no. 12.

4. R. Vaidyanathan et al., "Tongue-Movement Communication and Control Concept for Hands-Free Human-Machine Interfaces," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 37, no. 4, 2007, pp. 533–546.

5. M.-Z. Poh et al., "Heartphones: Sensor Earphones and Mobile Application for Non-Obtrusive Health Monitoring," *Proc. ACM Int'l Symp. Wearable Computers* (ISWC 09), 2009, pp. 153–154.

6. D. Matthies, "In-Ear BioFeed Controller: A Headset for Hands-Free and Eyes-Free Interaction with Mobile Devices," *Proc. ACM Conf. Computer–Human Interface* (CHI 13), 2013, pp. 1293–1298.

7. H. Manabe, M. Fukumoto, and T. Yagi, "Conductive Rubber Electrodes for Earphone-Based Eye Gesture Input Interface," *Proc. ACM Int'l Symp. Wearable Computers* (ISWC 13), 2013, pp. 33–40.
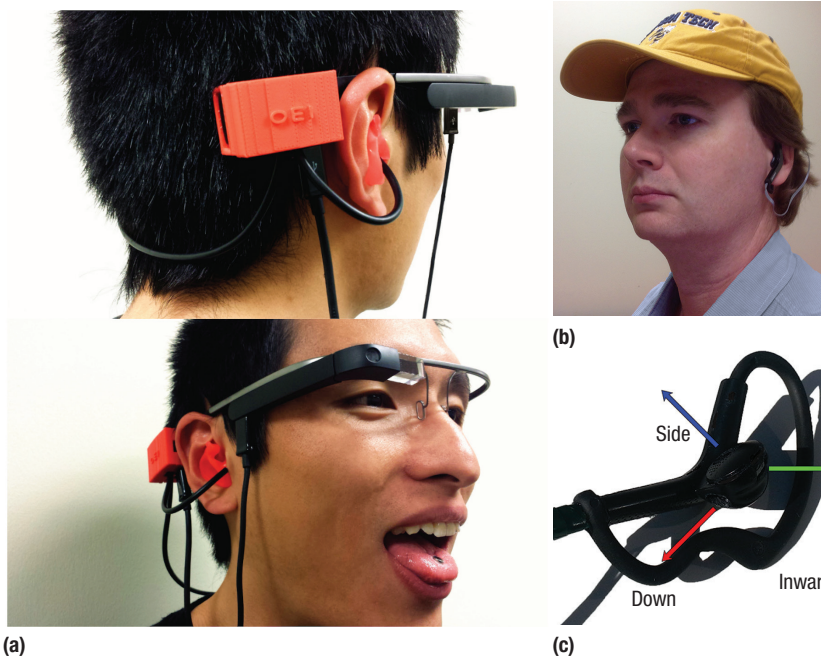
FIGURE 2. Prototype interfaces that integrate with Google Glass. (a) The tongue magnet interface (TMI) uses Google Glass's built-in magnetometer to sense a magnet affixed to the user's tongue, and the outer ear interface (OEI) control board fits over Glass's battery pod and connects to the proximity sensor embedded in an ear mold. (b) An improved OEI fits in an off-the-shelf Sony jogging headset. (c) The improved OEI has three orthogonal proximity sensors instead of one, which enables more accurate movement recognition.

data from the proximity sensors and pass it serially to a laptop for processing. Glass also sent its data to the laptop using its micro USB port.

As in the TDS experiment, each participant had a magnet temporarily glued to the tongue. Glass's magnetometer produced event timestamps that we recorded along with the sensor readings. We sampled the resulting five-dimensional datastream (three from the magnetometer and one each from the left- and right-ear proximity sensors) at 100 Hz.

Of the experiment's six subjects, three were native English speakers. We selected and labeled 11 phrases with 25 samples each. We dropped the "It hurts" phrase because we reasoned that its short length would result in poor performance. We again prompted subjects with a randomized phrase sequence on a computer screen in regular intervals, but in this experiment, breaks occurred every 30 samples instead of every 10.

At each prompt, the subject would silently speak the sample phrase in a natural manner. Again, the only restriction was to keep the head as steady as possible during recording. We collected each subject's data in one sitting. After verification, we discarded 24 phrase segments from the total 1,925 phrase samples collected because of sensor errors and data packet corruption.

As in the TDS experiment, we used HMMs for recognition, but in this experiment we added some refinements: for words with greater than two phonemes, we used a three-state, no-skip topology, and for words with two or fewer phonemes, we allowed a skip state (1→3).

We constructed phrase HMMs by stringing together each word topology. For example, we represented the phrase "Give me my medicine, please" with a 15-state HMM with two skips— (4→6) and (7→9)—for the words "me" and "my."

To model each state, we used a mixture of Gaussians (MoG). In MoG HMMs, the learned probability distribution function to represent each state is a weighted sum of Gaussians. This technique allows more freedom in approximating probability distribution functions. We fixed the number of Gaussians at two and, as in the feasibility test, used diagonal covariance matrices to represent all states. We added start and end states to model the pause before and after the phrase was uttered.

When using all magnetometer and sensor features, the average user-dependent recognition accuracy for the six subjects was 90.5 percent, and the highest single-subject accuracy was 96 percent. The greatest confusion occurred between the phrases "The medicine hurt me" and "Give me my medicine, please." Recognition is likely to improve with more careful phrase selection.

Overall, our results are encouraging. With more training data, we should be able to improve system performance and incorporate more phrases.

EXPERIMENT WITH THE OEI ALONE

As part of our goal to make silent-speech recognition systems truly nonintrusive and easy to use, we conducted a phrase-recognition experiment with the OEI alone. Using the OEI alone would eliminate the need for tongue piercing as a permanent solution, which would raise the user-acceptance level considerably.

Data collection and results

We used OEI data from one subject to train the HMMs and quickly discovered that the TMI and OEI did not easily recognize the same phrases. For example, for the phrase "Help me, I need assistance," the subject in the TMI + OEI experiment showed a false-positive rate of 10 percent, while the same subject had a false-positive rate of only

FIGURE 3. How the OEI works. (a) The OEI sits at the outer ear canal between the mastoid bone and the mandibular condyle. The proximity sensor's infrared signal (red) reflects off the ear canal (blue) back to the sensor's photocell, gradually getting weaker as it leaves the point of reflection. (b) When the mouth opens, the mandibular condyle slides forward and creates a small void. The tissue surrounding the ear canal fills this void, changing the shape of the ear canal. The user's jaw movement from silent speech is reflected in the ear canal's deformation and, correspondingly, in the changed infrared reflection sensed by the proximity sensor. These changes through time allow the OEI to recognize silently spoken phrases.

3 percent with the OEI alone. These results imply the need to carefully design the phrase set to be distinct and easily recognizable to both individual sensors and sensor combinations.

Recognition accuracy was highest (85.34 percent) when we reduced the phrase set from 11 to 9 phrases, deleting "Help me, I need assistance" and "The medicine hurt me." We suspect that designing phrases more carefully to induce jaw movement would improve accuracy.

We also investigated whether the system with the OEI alone could recognize the phrase "Attention, Google Glass." We collected one hour of data for a single subject using the OEI with one proximity sensor. The system was able to classify 17 of the 18 true positives and triggered falsely five times.

Although these results show promise, clearly much work remained to make a system with the OEI alone a reasonable product. To address that need, we focused the rest of our OEI experiments on improved versions of OEI hardware.

Sensor modifications

The structure and appearance of the human ear varies among individuals.[5] Consequently, the use of off-the-shelf ear molds might not be the best way to place OEI sensors. Even custom ear molds can become unstable because of ear-canal deformation, particularly when the user must perform mouth gestures that require a wide range of movement, such as sliding the lower jaw left and right. This movement can cause the ear mold to drop or shift from its initial location, resulting in a partial or complete block of the sensor's signal.
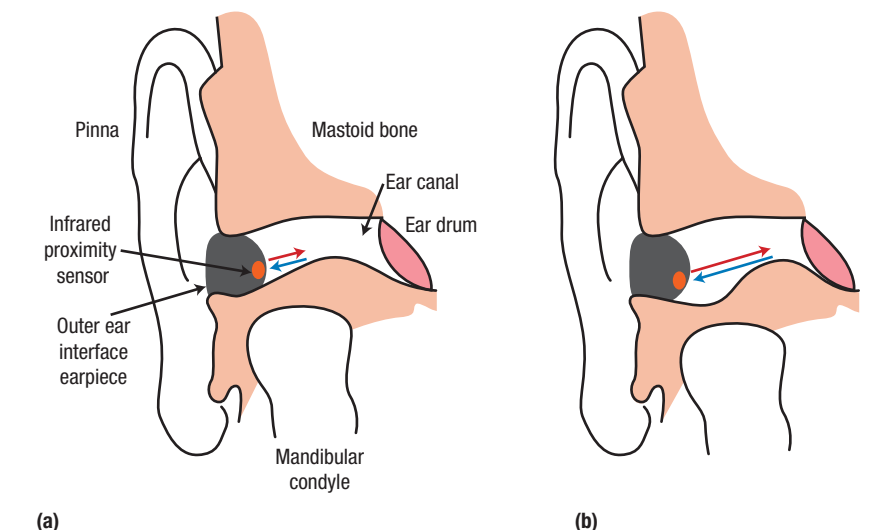
To overcome these issues, we replaced ear molds with Sony sports earphones, which have a stabilizing and fitting mechanism that is unaffected by ear-canal deformation (see Figure 2b). The earpiece has a loop around the outer ear that secures its placement. The loop length is adjustable to suit a wide range of ear sizes.

We added two more proximity sensors in the earpiece, placing each sensor orthogonally to the other two (see Figure 2c). This configuration provides wider coverage of ear-canal deformation and provides additional input channels if one of the sensors is blocked.

EXPERIMENT WITH IMPROVED OEI AND SIMPLE JAW GESTURE

The OEI is also useful in recognizing intentional jaw gestures. For example, when Glass makes a sound alerting the user to an incoming message, a simple left-to-right jaw movement might wake Glass's monitor to display the message. We conducted a fourth experiment to evaluate the feasibility of using the OEI in this way.

We collected data from 22 subjects while they performed a simple left-to-right jaw gesture, sat at rest, ate, drank, looked around the room, and talked. We manually segmented each subject's gesture data into 2.5-second segments, sampled segments of the same length from false-positive scenarios, and included those segments in the dataset.

We passed each segment through a bandpass finite-impulse-response filter with cutoff frequencies of 0.5 and 2.0 Hz—a time scale that corresponded with the jaw gestures. We then subtracted the mean from each axis and normalized segments between 0 and 1. We used a five-second window of resting data to construct a principal component axis (PCA) projection matrix for each subject and projected the remaining data segments from a subject onto this axis to normalize the data. Because the greatest variation during resting is the heartbeat pulse in the ear canal, we reasoned that PCA preprocessing should align each user's data along that axis, which we hoped would also align with the deformation of the ear canal during the left-to-right jaw gesture.

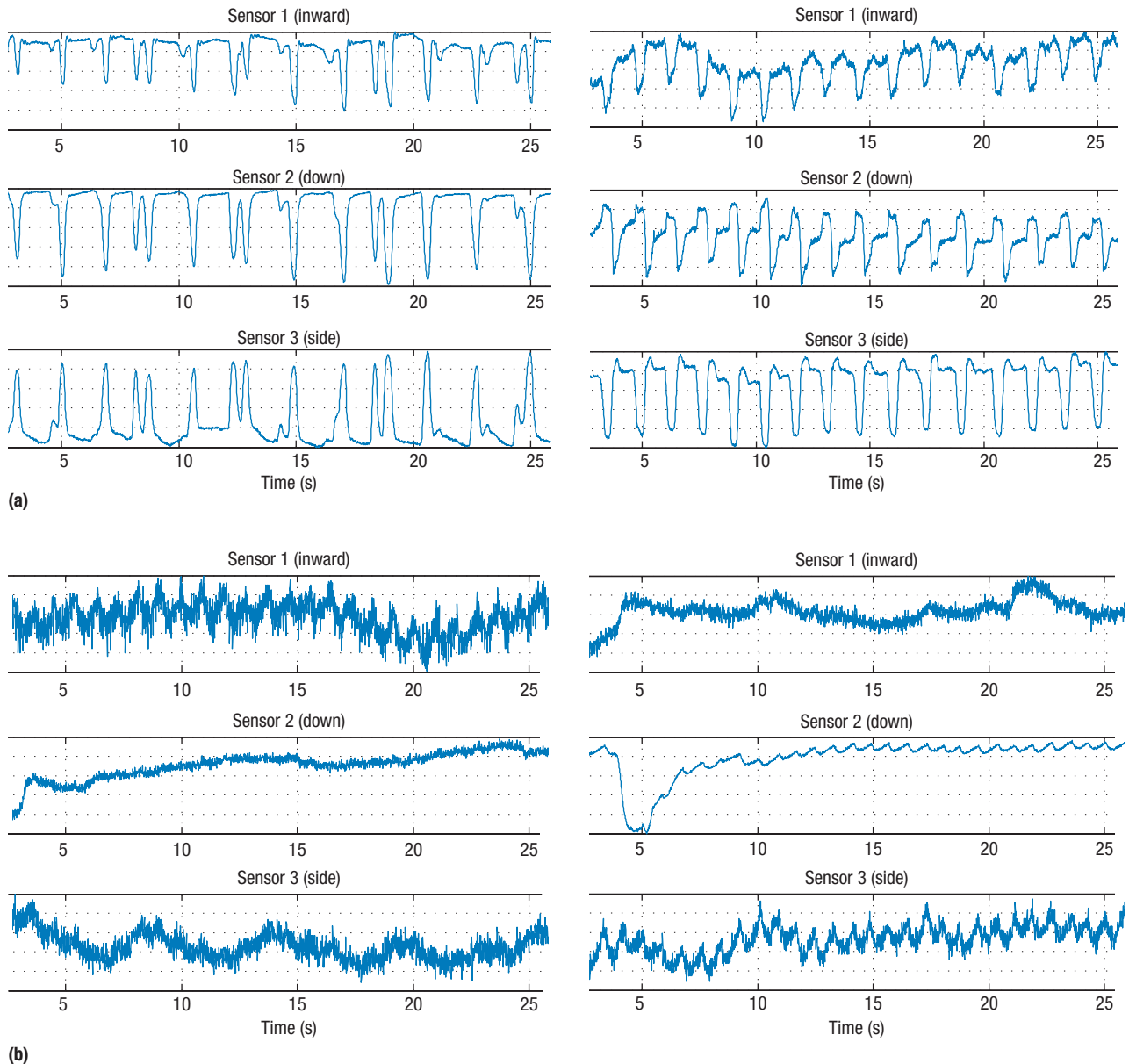We used a dynamic time warping (DTW) distance metric to classify data

FIGURE 4. Normalized data from the improved OEI system. (a) Comparison of a left–right jaw gesture made by two different participants (left and right) showing the variation in signals that can be expected. (b) The users' resting heartbeat signals. Although the heartbeat signal is not nearly as strong as the jaw motion, it can be useful in calibrating the OEI's proximity sernsors. The left user's heartbeat registers best in the inward sensor, whereas the right user's heartbeat is strongest in the side sensors.

using k nearest neighbors. We restricted the warp path within a Sakoe-Chiba band of 50 samples.[6]

We performed user-independent cross-validation by omitting one subject from the training set. The recognition results were averaged across all users. Experiments were also performed in a user-dependent manner with an 80/20 percent training/testing split. We used a 10-fold Monte Carlo cross-validation scheme, and the results averaged over the folds.

In the user-independent experiment, the system achieved an 84.51 percent overall classification accuracy, although it failed to identify a little more than a quarter of the jaw gestures (10.33 percent false positives and 27.5 percent false negatives). The user-dependent experiment yielded much better results: 97.64 percent overall accuracy, 2.6 percent false positives, and 1.7 percent false negatives.

To train the classifier in the user-dependent experiment, we used 24 seconds of samples of each daily activity and 48 seconds of jaw-gesture examples.

ABOUT THE AUTHORS

ABDELKAREEM BEDRI is an MSc student in the College of Computing at Georgia Tech. His research interests include gesture recognition, wearable and interactive systems, and assistive technologies. Bedri received a master's by research in systems engineering from the University of Reading, UK. Contact him at abedri6@gatech.edu.

HIMANSHU SAHNI is a PhD student in the College of Computing at Georgia Tech. His research interests include interactive machine learning, reinforcement learning, and computer vision. Sahni received an MS in computer science from Georgia Tech. Contact him at himanshu@gatech.edu.

PAVLEEN THUKRAL is a computer science undergraduate student at Georgia Tech. His research interests include pattern recognition, computer vision, and sparse-time-series recognition tasks. Contact him at pav920@gatech.edu.

THAD STARNER is a professor of computing at Georgia Tech. His research interests include pattern discovery, wearable computing, and sign language. Starner received a PhD in wearable computing from MIT. He is a member of IEEE and ACM. Contact him at thad@gatech.edu.

DAVID BYRD is a research scientist in the Interactive Media Technology Center (IMTC) at Georgia Tech. His research interests include machine learning, complex data analysis, pattern recognition, supervised learning for financial markets, and data visualization. Byrd received a BS in computer science from Georgia Tech. Contact him at db@gatech.edu.

PETER PRESTI is a senior research scientist in the IMTC and codirector of the Wearable Computing Center at Georgia Tech. His research interests include wearable and embedded applications, pattern recognition, mobile electronics, and computer vision. Presti received an MS in computer science from Georgia Tech. Contact him at peter.presti@imtc.gatech.edu.

GABRIEL REYES is a PhD student in computer science and human–computer interaction in the School of Interactive Computing at Georgia Tech. His research interests include embedded hardware and novel sensing for wearable input/output systems, gestural interactions, and activity recognition. Reyes received an MS in electrical engineering and an MS in international business from the University of Florida. Contact him at greyes@gatech.edu.

MAYSAM GHOVANLOO is an associate professor of electrical and computer engineering and founding director of the Bionics Laboratory in the School of Electrical and Computer Engineering at Georgia Tech. His research interests include implantable microelectronic devices, neuroprosthetics, assistive technologies, smart health, and medical instrumentation. Ghovanloo received a PhD in electrical engineering from the University of Michigan, Ann Arbor. He is a senior member of IEEE and an associate editor of *IEEE Transactions on Biomedical Circuits and Systems* and *IEEE Transactions on Biomedical Engineering*. Contact him at mgh@gatech.edu.

ZEHUA GUO is an electrical engineer in the Surface team at Microsoft. His research interests include wearable technologies, analog integrated circuits, and sensor networks. Guo received a BS in electrical engineering from Georgia Tech. Contact him at zeguo@microsoft.com.

Because jaw-gesture training requires little active effort, it might be reasonable for users to train the system to recognize their particular jaw patterns in practice.

HEARTBEAT AS A CALIBRATION SIGNAL

Heartbeat monitoring can help a wearable computer know when the sensor is currently in the user's ear and working properly. If the proximity sensors' signal variance is too high, the system knows that the user is not wearing the earpiece and can turn off recognition. If the proximity readings are appropriate but there is no signal variability from the heartbeat, the system knows not only that the user is not wearing the earpiece but that it is some place where the proximity sensors are covered, perhaps in a drawer.

Figure 4 shows normalized data from the improved OEI system with its three proximity sensors. Figure 4a shows that the shape of a jaw gesture in the data can vary significantly across users. In addition, different users show the strongest signal in different sensors. However, we might be able to predict which sensor, or combination of sensors, will provide good detection of the jaw gesture by observing the user's heartbeat in the signal while the user is at rest. Figure 4b shows that the heartbeat is strongest in the inward sensor for the left user, which—before

normalization—corresponded to the sensor with the strongest signal for the jaw gesture as well. Similarly, for the right user, the heartbeat and jaw gesture are strongest in the side and down sensors. Furthermore, by looking at the strength of the heartbeat signal, the system might predict the strength of the jaw gesture signal and ignore other signals that are too large or too small.

Surprisingly, for many participants, the downward-facing sensor shows a stronger signal than the sensors pointing into the ear canal. This phenomenon is caused by the facial muscles which change the shape of the ear's exterior as they move. By using PCA, we can automatically condition the data to take advantage of the best signal. In fact, such a preprocessing step has proved beneficial in preliminary tests for creating user-independent recognizers.

Our results suggest that sensing the motion of a single passive magnet mounted on a user's tongue with one headset-mounted magnetometer, combined with a proximity sensor in the ear, is sufficient to recognize a preselected set of 11 silently spoken phrases with high accuracy given user-dependent training.

Promising preliminary work also shows that a set of three proximity sensors mounted in an earphone might allow a user to control a wearable device through jaw gestures. Such a system would be more acceptable to the average consumer than a tongue magnet and could provide other benefits such as sensing the user's resting heart rate.

Performance using silent speech and jaw gestures could significantly improve with the addition of a microphone and gyroscope to the earphone, which would help distinguish head motion, eating, drinking, and speech from gestures of interest. We are currently investigating the feasibility of such an integrated earphone system, which could control Google Glass or another mobile device though relatively subtle jaw movements, yet would differ only marginally from headphones now on the market. ⊏

REFERENCES

1. H. Sahni et al., "The Tongue and Ear Interface: A Wearable System for Silent Speech Recognition," *Proc. ACM Int'l Symp. Wearable Computers* (ISWC 14), 2014, pp. 47–54.
2. M. Ghovanloo, "Tongue Operated Assistive Technologies," *Proc. 29th Ann. Int'l Conf. IEEE Eng. in Medicine and Biology Society* (EMBS 07), 2007, pp. 4376–4379.
3. S. Young et al., "The HTK Book (for HTK version 3.4)," tech report, Cambridge Univ. Eng. Dept., 2006.
4. S. Darkner et al., "Analysis of Surfaces Using Constrained Regression Models," *Proc. Conf. Medical Image Computing and Computer-Assisted Intervention* (MICCAI 08), 2008, pp. 842–849.
5. D.J. Hurley, A.-Z. Banafshe, and M.S. Nixon, "The Ear as a Biometric," *Handbook of Biometrics*, A.K. Jain, P. Flynn, and A.A. Ross, eds., Springer, 2008, pp. 131–150.
6. H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, 1978, pp. 43–49.